Kwong Jack (2012) “Resisting Aliefs: Gendler on Belief-Discordant Behaviors”, **Philosophical Psychology** (Feb 2012) vol. 25, no. 1, pp. 77-91. [ISSN: 0951-5089] The version of record is published by Taylor & Francis available online: http://dx.doi.org/10.1080/09515089.2011.569915

Resisting Aliefs: Gendler on Belief-Discordant Behaviors

Jack M. C. Kwong

Abstract: This paper challenges T. S. Gendler’s notion of aliefs, a novel kind of mental state which she introduces to explain a wide variety of belief-discordant behaviors. In particular, I argue that many of the cases which she uses to motivate such a mental state can be fully explained by accounts that make use only of commonplace attitudes such as beliefs and desires.

1. Introduction

T. S. Gendler has recently introduced the concept of ‘‘alief’’ to explain a wide variety of belief-discordant behaviors (Gendler, 2008a, 2008b). According to her, these are cases in which people exhibit behaviors that are prima facie difficult to reconcile with what they explicitly believe and desire. Gendler argues that the best way to explain these cases is to incorporate the notion of ‘‘aliefs,’’ which can be characterized roughly as mental states—typically ones with representational, affective and behavioral components—that directly give rise to a set of response patterns automatically triggered by some stimulus in a person’s environment. The reason why there is a mismatch in belief and behavior, Gendler proposes, is that these people entertain and act on an alief with a content that runs counter to their belief. Moreover, she argues that any account of such behaviors that makes use of beliefs and desires needs to make room for aliefs in order to be explanatorily adequate. In this paper, I argue that many of the belief-discordant cases which she uses to motivate the notion of aliefs and to illustrate the inadequacy of the belief-desire accounts can be fully explained without recourse to these novel mental states. In particular, I show that theories that make use of commonplace mental states such as beliefs and desires already have the resources to explain the belief-behavior mismatch.

2. The Nature of Aliefs

To motivate the notion of alief, Gendler argues that it can be used to explain a cluster of ‘‘baffling’’ cases that involve a mismatch between belief and behavior. For example, consider the Skywalk of the Grand Canyon, a horseshoe shaped glass walkway built in 2004 that suspends 4000 feet above the Colorado River. Although visitors believe that the walkway is safe and stable, they do not behave in a way that is appropriate to their belief. Instead of eagerly trotting down towards the centre of the Skywalk for a better view, many are reluctant to walk across it. Their behavior is thus discordant with their belief. Another example Gendler discusses comes from one of Paul Rozin's experiments, which demonstrates that people are averse to eating soup served from brand new bedpans (Haidt, McCauley, & Rozin, 1994; Rozin, Haidt, & McCauley, 2000; Rozin, Millman, & Nemeroff, 1986). These people clearly believe that the bedpans are new and sterile. Yet, this belief is not accompanied by appropriate behavior. Their reluctance is, again, considered to be a belief-discordant behavior.

How do we explain these belief-discordant behaviors and, in particular, what is the best way to describe the cognitive state of people who exhibit such behaviors? Gendler argues that an adequate description cannot be given in terms of what she calls the ‘‘classic cognitivist picture,’’ which explains these cases by appealing to traditional folk-psychological categories such as false beliefs, forgetfulness, selfdeception, vivid imagination, superstition, hypocrisy, and doubt. These explanations are inadequate because they all turn out to be false, misleading or incomplete (Gendler, 2008b, p. 561). For example, to describe the Grand Canyon visitor as having temporarily forgotten that the Skywalk is safe seems to have a limited application, for she would still manifest the same reluctance even if she fully and consciously entertained the thought of the structure's safety, ‘‘with explicit attention to its meaning and implications’’ (Gendler, 2008a, p. 640). Likewise, to say that the visitor has doubts about the structure of the glass walkway is to misdescribe the situation: if she were not completely certain that the Skywalk was safe, she would, most likely, not even contemplate taking a step on the glass platform in the first place.

In order to provide an adequate explanation of these instances, Gendler proposes that we incorporate the notion of alief. Here is how she provisionally characterizes it:

> To have an alief is, to a reasonable approximation, to have an innate or habitual propensity to respond to an apparent stimulus in a particular way. It is to be in a mental state that is. . . associative, automatic and arational. As a class, aliefs are states what we share with non-human animals; they are developmentally and conceptually antecedent to other cognitive attitudes that the creature may go on to develop. Typically, they are also affect-laden and action-generating. (2008b, p. 557)

Also:

> A paradegmatic alief is a mental state with associatively linked content that is representational, affective and behavioral, and that is activated—consciously or nonconsciously—by features of the subject's internal or ambient environment. Aliefs may be either occurrent or dispositional. (Gendler, 2008a, p. 642)

To have an alief, then, is to be in a mental state in which certain response patterns, typically with a representational, affective and behavioral component, are automatically activated by some stimulus in a person's environment. To illustrate, in the case of the Skywalk, the sight of the surroundings (the stimulus) automatically triggers in

the visitor a response pattern, the content of which consists of the visual impact of the deep chasm (the representational component), feelings of fear and panic (affective), and an activation of motor routines for retreat (behavior). Put another way, the sight of the chasm renders occurrent in the visitor the alief with the content ‘‘really high up, long long way down. Not a safe place to be! Get off!!’’ (Gendler, 2008a, p. 635). According to Gendler, the reason the visitor expresses reluctance in walking across the Skywalk is that this alief is activated alongside her belief that the walkway is structurally secure. Although her belief represents the situation as being safe, her alief, which is activated without the aid of conscious thought, depicts the situation as a dangerous one and in so doing, makes her feel anxious and prepares her to retreat from the departure platform. Her behavior is discordant with her belief because it reflects an alief with a contrary content.

As mentioned, Gendler’s characterizations are provisional. For instance, she mentions the possibility that some aliefs, unlike the paradigmatic cases, may ‘‘involve the mental activation of a different sort of associative cluster’’ (2008a, pp. 643–644) they may lack a representational, affective or behavioral component, or may involve associations that cannot be ‘‘easily subsumed under [these] three categories’’ (2008a, pp. 643–644). Another issue on which she is tentative is whether it might be better to think of an alief as a two-place (S alieves R) or a four-place relation (S alieves R-A-B; Gendler, 2008a, p. 645). To be sure, these are all important considerations. However, my main goal in this paper is not to criticize the nature of aliefs, but to show that they are not required to explain the above cases of belief-discordant behaviors. I therefore propose to leave aside, for now, some of these considerations, and concentrate only on those that are directly relevant to my argument below. Two are particularly pertinent.

First, aliefs are reality-insensitive in that they can occur even in the face of contrary evidence. Thus, despite the assurance that the walkway is safe, the visitor still alieves it to be dangerous. One of the reasons why aliefs are not sensitive to evidence is that they reflect ‘‘associational patterns that have been laid down in our minds as the result of our experiences and those of our genetic ancestors’’ (Gendler, 2008a, p. 651). As such, they are triggered automatically in the presence of the relevant stimulus, even if the stimulus is merely apparent and does not reflect how reality is. Such insensitivity highlights a fundamental difference between alief and belief. Whereas the latter aims to track truth in the world and responds directly to evidence, the former does not. Consequently, aliefs violate norms related to cognitivebehavioral coherence. Second, since aliefs are laid down in our minds in the ways just mentioned, they are difficult to regulate. In order to bring aliefs in line with our reflective commitments, Gendler suggests that we either cultivate belief-concordant habits by consciously behaving in ways that are consistent with our reflective commitments, or employ our imaginative resources to reshape our behavioral dispositions so that they correspond with our beliefs (2008b, section 4).

3. Reviving the Classic Cognitivist Picture

As noted, Gendler argues that the classic cognitivist picture is unable to explain adequately the aforementioned instances of belief-discordant behaviors. Indeed, she claims that ‘‘any theory that makes appeal to notions like belief, desire and pretense in order to explain behavior needs to make appeal to (something like) alief in order to make sense of a wide range of otherwise perplexing phenomena’’ (Gendler, 2008b, p. 555). Elsewhere, she notes that ‘‘if you want to take seriously how human minds

really work, and you want to save belief, then you need to make conceptual room for the notion of alief'' (Gendler, 2008a, p. 642). In this section, I argue that many of the cases of belief-discordant behaviors, which Gendler uses to motivate the notion of alief and to expose the shortcomings of the classic cognitivist picture, can indeed be adequately explained without recourse to aliefs. Specifically, my target will be the sort of belief-discordant behaviors which she discusses in the opening paragraphs of ''Alief and belief'' (Gendler, 2008a). Here, she introduces and characterizes this novel type of mental state, shows how it can make sense of belief-discordant behaviors, and explains how aliefs differ from beliefs, states of imagination and other attitudes (Gendler, 2008a, pp. 634–637, sections 1 & 2).[1] In my view, the key to explaining these behaviors without appealing to the notion of alief is to broaden our consideration of the subject's beliefs beyond those identified by Gendler. Of these beliefs, I argue that one set in particular plays a critical role in explaining the mismatch between the subject's belief and behavior.

Two caveats are in order. First, it is important to note that the examples of beliefdiscordant behaviors with which I am concerned in this paper do not exhaust the cases which Gendler uses to discuss the notion of alief. She also refers to other examples, including those of ''automaticity'' and animal behaviors (e.g., frogs and beebees, puppies and their own mirror images; Gendler, 2008b, pp. 552–553; Gendler, 2008a, section 3). In the following, whenever I use the expression ''beliefdiscordant behaviors,'' I shall be referring only to the examples specified in the previous paragraph. The main reason for omitting ''automaticity'' cases and animal behaviors is that the failure of the classic cognitivist picture to adequately explain them, in my view, does not have the same ramifications as its failure to explain the ones with which I am concerned. For instance, it is not entirely clear whether the classic cognitivist account was designed to explain behaviors resulting from subliminal priming or cases relating to animals. I shall return briefly to this issue in the final section of this paper. For now, suffice it to note that the kinds of behavior to be considered from the standpoint of my argument below are hereby identified. Second, for convenience of exposition, the following discussion will revolve solely around Rozin's experiment involving a piece of fudge that is shaped like dog feces, which is one of the cases used by Gendler to motivate the notion of aliefs. At certain junctures, I will modify the experiment slightly so as to provide a more comprehensive description of the subject's cognitive state, and to pave the way for the kind of description that can be extended and applied to other belief-discordant behaviors that Gendler has considered. I will show how this description can be generalized to adequately explain all the cases of belief-discordant behaviors mentioned above.

Rozin's experiment is as follows: subjects were initially offered a piece of highquality fudge in the shape of a square, which they ate without hesitation. They were then presented with two more pieces of fudge, one of which was shaped like a disc and the other like dog feces. Moreover, they were told that both of these pieces came from the same source as the square-shaped fudge, which they had just enjoyed. Rozin found that subjects were highly reluctant to eat the fudge that was shaped like dog feces (Rozin et al., 1986). According to Gendler, this counts as a belief-discordant behavior. As she notes, these subjects ''believe that the fudge has not changed its chemical composition'' and, if asked directly, would ''show no hesitation in endorsing [it]'' (Gendler, 2008a, p. 636). Yet, their belief did not lead to appropriate behavior. Instead of cheerfully eating that piece of fudge, they were reluctant to do so. Gendler's explanation is that the sight of the dog feces-shaped fudge rendered occurrent in these subjects, alongside their belief, the alief with the content ''filthy

object! Contaminated! Stay away!'' (2008a, p. 636). Their behavior, in other words, reflected an alief whose content is contrary to their belief.

Gendler's explanation appeals to an additional kind of mental state, which the subject holds alongside her belief.[2] However, an alternative explanation is available if we simply explore other beliefs that the subject may already hold. Following Gendler, we start by attributing to the subject the explicit belief with which her behavior is discordant,

> (a) This piece of fudge (in front of me) is composed of a substance that I consider delicious and appealing.

To an approximation, it also seems natural, given the setup of the experiment, to attribute to her the following beliefs:

> (b) This piece of fudge (in front of me) looks like dog feces.
> (c) Any food item that looks like dog feces tends to be unappetizing.[3]
> (d) Any food item that is unappetizing should generally be avoided.[4]

In my view, these are beliefs that can be plausibly attributed to the subject. Briefly, she believes (b) because the piece of fudge presented to her has been deliberately shaped like dog feces and intended to strike her as such. She believes (c) because food items that look like dog feces remind her of the real thing, which she finds to be disgusting. Finally, she believes (d) because it has practical consequences: at the very least, avoiding unappetizing things helps to prevent her from experiencing discomfort or disgust. Additional support for attributing these beliefs to her may be obtained by employing the direct but fallible procedure of asking her whether she holds these beliefs and by examining her answers in light of her personal history. As with her belief (a), she would very likely, I suspect, endorse these claims.[5]

Admittedly, (b)–(d) are approximations and may not be the exact beliefs that the subject holds. Instead of holding a belief specifically about dog feces, she may think that the object in front of her looks like nondescript feces. Nor are these beliefs the only ones that she possibly entertains. Knowing that the dog feces-shaped object is in fact fudge, she may also believe that it is edible, appealing and/or delicious. These are important considerations to which I will return shortly. For now, I will continue with the argument.

If it is plausible to attribute the belief set (b)–(d) to the subject, it is also plausible to attribute the belief that it supports, namely,

> (e) This piece of fudge should be avoided.

We are now in a position to explain the subject's cognitive state in the experiment. To adapt from Gendler's explanation, we can maintain the following: the subject surely believes that the fudge has not changed its chemical composition. But alongside this belief there is something else going on in her state of mind. Although she believes that the piece of dog-feces shaped fudge is harmless, she also believes something very different: she believes (b)–(e). To elaborate, when the subject is presented with a piece of fudge that is shaped like dog feces, not only does she explicitly believe that it is a piece of fudge, she also believes, at the same time, that it looks like dog feces. Put another way, the sight of the dog feces-shaped piece of fudge brings about her belief (b). Since she also believes that food items that look like dog feces are—due to their appearance—unappetizing, and that unappetizing food should generally be avoided, she thus takes the appropriate measure to avoid eating that piece of fudge. Depending on the relative strengths of her relevant desires—that

is, her desire to avoid unappetizing food and her desire to eat delicious fudge—she may either actively refuse to eat the fudge (in which case the fudge's appearance is simply too sickening), or she may eat it but do so with a residual sense of discomfort or disgust.[6] Either way, the subject's behavior reflects something other than her belief (a): it reflects beliefs (b)–(e), and the corresponding desire.

This account, in my view, constitutes an adequate explanation of the subject's cognitive-behavioral state. More importantly, it does so without recourse to aliefs, relying instead primarily on beliefs and desires. The key here is to explain how it is that the subject comes to entertain a content that runs counter to her explicit belief, and continues to do so in the face of available evidence that supports her belief. Gendler's proposal is, of course, that the sight of the dog feces-shaped piece of fudge automatically activates in the subject the alief with the content ''filthy object! Contaminated! Stay away!'' But that is not the only possible way for the subject to entertain a mental state with a contrary content. As suggested, she could also be having parallel beliefs about what the piece of fudge looks like and what it really is. More importantly, so long as the piece of fudge looks like dog feces, she will continue to hold (b) regardless of how much evidence she is given about it as tasty fudge. In other words, she will hold onto her belief that it looks like dog feces even if she is presented with a complete chemical analysis of its ingredients, or is reminded (possibly through a demonstration) how the high-quality fudge came to be shaped like dog feces. Indeed, she will continue to do so after she eats the fudge and declares how delicious it is! What this shows is that her belief (b), like Gendler's alief, is responsive to apparent stimuli. And once the subject responds to such a stimulus, she will come to entertain additional beliefs that are relevant to (b) and behave, in the way explained above, in a belief-discordant manner.

At this point, a worry arises. As mentioned, Gendler argues that aliefs are insensitive to evidence, and thus violate certain cognitive-behavioral norms; this in turn disqualifies them from being beliefs. To elaborate, it is instructive to examine a problem which Gendler raises against certain explanations of belief attribution offered by the classic cognitivist picture.[7] According to her, one attempt to explain belief-discordant behaviors is to argue that the subject's competing tendencies can be used to credit her with different beliefs (Gendler, 2008b, p. 562). In the case of the dog feces-shaped piece of fudge, the subject's tendency to express reluctance in eating the fudge, on the one hand, can be used to credit her with the belief that the piece of fudge has somehow changed its composition (or that it is no longer appealing to eat). On the other hand, her tendency to eat the fudge can be used to credit her with the belief that the piece of fudge has not changed its chemical composition (and thus, is appealing to eat). Both of these beliefs can then be used to describe her cognitive state: her reluctance to eat the fudge is due to the concurrent presence of contesting beliefs in her mind.

The problem with this explanation, Gendler argues, is that the classic cognitivist picture treats these competing tendencies as ''being on a par,'' each being ''sufficient to credit the subject with the relevant belief'' (2008b, p. 563). In her view, these tendencies are not on a par, and in order for an attitude to count as a genuine belief, it must at least track truth in the sense that it is ''subject to immediate revision in the face of changes in our all-things-considered evidence'' (Gendler, 2008b, p. 565). Notice that only one of the subject's above-mentioned tendencies tracks truth in this sense: the subject would immediately revise her belief that the fudge has not changed its chemical composition if she were presented with evidence to that effect. Her other putative belief about the shape of the fudge, however, would not go away: she would

continue to hold the attitude that the piece of fudge has somehow changed its composition (or that it is no longer appealing to eat) even in light of evidence to the contrary. Consequently, the explanation fails.

To advance my point, the relevant issue to consider next is whether (b) tracks truth in Gendler's sense. The answer is yes. It does so in that the subject would immediately revise (b) if the piece of fudge were molded into a different shape, say, the shape of a muffin. Accordingly, she would then believe that it looks like a muffin. Notice that (b) is only sensitive to a certain kind of evidence, namely, that which has a direct bearing on the appearance of the piece of fudge. This explains why she need not revise or abandon (b) even if she were presented with the result of a full chemical analysis, which does not have much to do with how the fudge looks. Indeed, (b) would still be tracking truth about the appearance of the dog feces-shaped object even if the object is proven to be high-quality fudge. Since (b) respects the normative constraint of tracking truth, it qualifies, at least by this standard, as a belief.

Incidentally, the subject's other attitudes, (c) and (d), also respect this normative constraint, albeit with one important difference. Qua generalizations, they do not, unlike (b), need to be subjected to immediate revision in the face of contrary evidence. To illustrate, suppose the subject reluctantly eats the dog feces-shaped fudge and finds it to be surprisingly delicious. She will then have evidence that not all unappetizing things should be avoided.[8] But having this piece of evidence may not be sufficient for her to revise her belief (d), that anything unappetizing should generally be avoided. This is because one counterexample may not be sufficient to falsify or overturn the generalization. Only after she gets more and more into the practice of not wanting to avoid unappetizing objects will she begin to revise (d). Even then, she may only revise it partially, to something like ''anything that is unappetizing should more often than not be avoided.''

Two important points follow from this discussion. First, since (c) and (d) are subjected to revision in the presence of contrary evidence, they do not violate the normative constraint of tracking truth. As such, they, like (b), qualify as beliefs. Second, the way in which a person comes to revise generalizations such as (c) and (d) provides an explanation as to why belief-discordant behaviors, as Gendler has pointed out, are difficult to regulate. By my account, the person may first have to engage in the potentially difficult task of making these generalizations explicit, bringing them to the level of conscious thought, and then evaluating them in light of the evidence. In some cases, conscious deliberation may be required, as she will have to decide how much weight to give to a certain piece of evidence and to reflect on the sort of conditions under which she will overturn or modify her generalized beliefs. In other cases, she may simply have to wait for the right kind of evidence to show up. There is undoubtedly much more to be said about this complicated process of belief revision, but I hope to have at least shown how and why belief-discordant behaviors would also be difficult to regulate: it takes both time and conscious effort for the person to revise beliefs such as (c) and (d).

4. Objections and Replies

In this section, I consider three objections that may be raised against my argument.[9] The main strategy behind my account, as has been shown, is to broaden our consideration of the subject's beliefs beyond the ones which Gendler identifies to be

discordant with the subject's behavior. I have argued that the subject could plausibly be attributed an additional set of beliefs, consisting of a belief that is responsive to an apparent stimulus (belief (b)) and of beliefs with contents that are relevant to it (beliefs (c) and (d)). Together, these can adequately explain her belief-discordant behavior. One possible objection is that my account fails to consider certain beliefs that may also be attributed to the subject in addition to the belief set. For example, since the subject believes that the piece of fudge is composed of a delicious and appealing substance, it is reasonable to attribute to her the belief that:

(f) Anything that is delicious and appealing should generally be approached.

Together with (a), (f) would lead her to conclude that:

(g) This piece of fudge should be approached.

This presents a challenge to my account because the subject would now be attributed both the beliefs that the fudge should be avoided and that it should not be avoided.[10]

In response, I agree that it is plausible to attribute (f) to the subject. Indeed, I suspect that many people probably consider fudge delicious and would not pass up the opportunity to eat some if given the chance. Still, (g) is not the necessary outcome, for the reason that having the belief (f) does not lead the subject to think that anything that is delicious and appealing must be approached. In other words, (f), a ceteris paribus generalization, allows for exceptions. In this context, I submit, it is reasonable for the subject to think that the repelling appearance of the oddly shaped fudge constitutes a sufficient reason for her not to approach it, even if it is composed of a delicious and appealing substance. Put differently, the piece of fudge, regardless of how delicious the subject may find it, can still be presented in a most unappetizing manner.[11] Her belief (f), then, may be captured as: ''anything that is delicious and appealing should generally be approached, unless it looks like something I find to be unappetizing, in which case I will avoid it.'' If this construal of (f) is correct, the subject would therefore not approach the fudge.

A related objection that can be raised is that my account seems to require that the subject holds precisely beliefs (a)–(e), whereas it is possible that she may not hold these exact beliefs. For instance, instead of (b), a belief specifically about dog feces, she may believe one of the following:

(b_0) This piece of fudge (in front of me) looks like nondescript feces.
(b_{00}) This piece of fudge (in front of me) looks like mud.
($b_{00\,0}$) This piece of fudge (in front of me) looks like non-fecal pile of muck.

Similarly, instead of (a), a belief about appeal and taste, she may believe one of the following:

(a_0) This piece of fudge is composed of a substance that is edible.
(a_{00}) This piece of fudge is composed of a substance that is safe to consume.

Any of these beliefs can be plausibly attributed to the subject. By the same token, it is easy to think of other alternatives, not only for (a) and (b), but also for (c), (d), and (e). Why then should we attribute (a)–(e) to her, as opposed to some other set of beliefs? There are, I think, two ways of addressing this objection. The first is that we could confirm our belief attribution by asking the subject whether she in fact holds them and, to the extent possible, by examining and evaluating her answers in light of

her overarching history. Thus, she may reveal that she indeed finds things that look

like dog feces unappetizing, and may recount experiences and reasons to support her claim. Conversely, we may discover that her admission is not supported by her history, in which case it would be appropriate to withdraw our attribution. This method of appealing to reflective verbal report is admittedly fallible and can be misleading. Nevertheless, it provides us with a prima facie basis for our attribution of beliefs (see footnote 5).

Another response to the objection is to point out that my argument allows for the possibility that beliefs other than (a)–(e) can be attributed to the subject. Suppose that the subject thinks that the piece of fudge looks like mud, which she finds to be particularly unappetizing. In this case, it would be appropriate to attribute to her the following beliefs:

(b_0) This piece of fudge looks like mud.
(c_0) Anything that looks like mud is unappetizing.
(d_0) Anything that is unappetizing should generally be avoided.

These beliefs in turn would lead her to conclude that this piece of fudge should be avoided. Since this belief set mirrors the structure of (b)–(d), it can be used in a similar way to explain the subject's belief-discordant behavior. The availability of alternative explanations, in my view, does not weaken my argument. On the contrary, it reinforces the central point of this paper, which is to show that theories that rely primarily on beliefs and other commonplace propositional attitudes do have the resources to explain belief-discordant behaviors without recourse to the notion of alief. All that is required is for the subject to hold a belief that is sensitive to an apparent stimulus and other beliefs that are relevant to such a (b)-like belief.

A third objection that can be raised against my argument is that it has a limited application. So far, I have developed my argument solely around the example of the piece of fudge that is shaped like dog feces. But what about the other cases of belief-discordant behaviors with which I am concerned in this paper? To what extent can my argument be applied to them? It can be used to explain all of them. Consider the aforementioned example of the subject who reluctantly walks across the Skywalk. The sight of the deep chasm under the glass walkway triggers in the subject the following beliefs:

(k) This glass walkway (in front of me) looks dangerous.
(l) Anything that looks dangerous tends to be scary.
(m) Anything that is scary should generally be avoided.

which, in turn, leads her to conclude that

(n) This glass walkway should be avoided.

The subject's reluctance to move towards the center of the Skywalk can then be explained by attributing to her beliefs (k)–(n) alongside her belief (j). Notice that this kind of explanation, with details adjusted to the specific scenario, can also be applied

to other cases of belief-discordant behaviors that are similar to the fudge and the Skywalk examples: people's reluctance to drink soup from brand new bedpans, to drink a glass of juice containing a sterilized dead cockroach, to place a piece of rubber in the shape of fake vomit between their lips, and so on. All that is required to explain each of these cases is for the subject to hold a set of beliefs that consists, most importantly, of a belief that is directly responsive to an apparent stimulus, and of beliefs that are related to such a (b)-like belief.[12]

5. Conclusion

The central aim of this paper has been to show that belief-discordant behaviors of the sort discussed above can be explained without employing the notion of alief. I have argued that theories that rely primarily on commonplace mental states such as beliefs and desires can sufficiently explain these behaviors. Moreover, they can do so without running into problems related to belief attribution and normative constraints. The strategy is to broaden our consideration to include additional beliefs to the one that is discordant with a behavior. In the above, I have identified a specific set of beliefs that can play a similar role to that of aliefs in explaining why there is a mismatch between belief and behavior. This set, which consists of a belief that is responsive to an apparent stimulus, and of beliefs related to such a (b)-like belief, explains, among other things, how it is that the subject comes to entertain a content that is contrary to her explicit belief, and why belief-discordant behaviors take time and effort to regulate. In terms of explaining the behaviors discussed in this paper, it is not necessary to invoke aliefs, and, contrary to Gendler's criticism, the classic cognitivist picture is not false, misleading or incomplete.

As mentioned earlier, this paper is only concerned with a specific set of beliefdiscordant behaviors that Gendler discusses. The reason for such a select focus is that these behaviors appear to be Gendler's primary references in formulating and characterizing the notion of alief and in discussing its explanatory powers. More importantly, they furnish the sort of evidence with which she argues that the classic cognitivist descriptions of belief-discordant behaviors are false, misleading and incomplete. What I have shown above is that, contrary to Gendler's appraisal of it, the classic cognitivist picture has indeed the resources to explain these behaviors adequately. But what about Gendler's other examples that fall outside the scope of my attention? What implications do they have for my argument? Towards the end of ''Belief and Alief'' (2008a), Gendler suggests that aliefs can also be used to explain cases involving ''automaticity''. These are cases in which behavior-inducing mental representations are ''activated by awakening the associative patterns that have come to be linked with some object, stereotype, protocol, or mental image'' (Gendler, 2008a, p. 656). Moreover, the resulting behaviors are induced without the mediation of any conscious processes. For example:

> Subjects performed a scrambled sentence task in which one group confronted sentences containing terms associated with the elderly (for example, wrinkle, bingo, and retired), whereas the second group's unscrambling task involved only neutral terms. After completing the experiment, subjects were surreptitiously timed as they walked down the hall to the elevator. Those primed with the elderly stereotype took significantly longer to walk to the elevator than those who had not been so primed. (Gendler, 2008, p. 659)

To explain the behavior of the second group, Gendler suggests that the terms related to the elderly caused in the subjects the alief with the content ''old. Tired. Be careful walking to that elevator,'' which in turn ''made them more likely to act in accord with it'' (2008, p. 659). At the present moment, I do not see how my account can be used to explain these types of cases involving automaticity; it would seem quite ad hoc and implausible to claim that subjects in these cases possessed some (b)-like belief when the majority of them did not even notice that they were exposed to terms associated with the elderly.[13] In regards to these ''automaticity'' cases, my view is that they perhaps merit consideration for introducing the notion of alief. Notice that this concession does not necessarily undermine the efficacy of the classic

cognitivist picture. My contention remains, briefly, that it is not entirely clear whether accounts that rely primarily on commonplace mental states such as beliefs and desires are in the business of explaining behaviors that result from subliminal priming manipulation.[14] If this view is correct, then in terms of the behaviors that such accounts are supposed to explain, their descriptions are adequate to the task.

But suppose that these behaviors do fall under the domain of these accounts. My contribution in this paper will then be more specifically the following. Given that many of the belief-discordant behaviors can be explained in terms of commonplace mental states such as beliefs and desires, Gendler seems to have overstated the range of ''perplexing phenomena'' in order to justify the need for aliefs. Actual cases that call for aliefs turn out to be far fewer than Gendler has suggested.[15] Moreover, though my account has only limited application, it nevertheless offers a basic defense of the classic cognitivist picture.

Acknowledgements

I would like to thank Chris Bartel, Brandon Conley, Jennifer Harker, Luke Kwong, and two anonymous referees for their insightful comments.

Notes

[1] I also have in mind the behaviors which she cites against the classic cognitivist picture in her ''Alief in action (and reaction)'' (Gendler, 2008b, pp. 560–571). Part of the reason for specifying these behaviors in such a vague way has to do with Gendler's overall strategy. As she notes, ''because alief is a novel notion, introduced to make sense of a cluster of otherwise baffling cases, most of ['Alief and Belief'] will proceed by examination of specific examples'' (Gendler, 2008a, p. 642).

[2] In the following discussion, I will refer to the subjects in Rozin's experiment in the singular.

[3] This belief is intended to capture the subject's subjective perspective on food items that (to her) look like dog feces. Similarly, belief (d) expresses what she would do in the presence of things she regards as unappetizing.

[4] The reason why it is stated that unappetizing food items should generally be avoided is that there are exceptions. For instance, guests at a dinner party may partake of unappetizing food in order not to offend their host. Similarly, people who are sick will typically consume unappetizing food for dietary reasons. In light of these exceptions, an alternative formulation of (d) could be something like, ''any food item that is unappetizing should—in the absence of social or health considerations of the sort considered above—be avoided or passed on.''

[5] Here, I intend to give these reflective verbal reports the same weight as Gendler does when she uses them to establish the subject's belief-discordant behavior. For example, in the case of the Skywalk, Gendler asserts that it seems clear that the subject believes that the walkway is safe: ''ask the subject directly and she will show no hesitation in endorsing such [a claim] as true. Ask her to bet, and this is where she will place her money. Ask her to think about what her other beliefs imply and this is what she will conclude. Look at her overarching behavior and this is what it will point to'' (2008a, p. 638).

[6] She experiences a residual sense of disgust or discomfort because she would be eating the piece of fudge while entertaining thoughts about dog feces (namely, beliefs (b) and (c)).

[7] This discussion is adapted from Gendler's discussion of precipice cases (2008b,

pp. 561–566).

[8] It is important to note that her experience with this particular piece of fudge may not be sufficient for her to conclude that dog feces-shaped fudge as a kind of object should not be avoided; she may very well regard it as an exception. My view is that only after she has encountered more and more instances of delicious dog feces-shaped fudge will she begin to revise her belief (d).

[9] I am indebted to the journal's referees for pointing out some of these objections.

[10] Alternatively, with the inclusion of these additional beliefs, the subject may simply conclude that (g) the piece of fudge should be approached. There are two ways to respond to this challenge. The first is to agree that she may very well make such a conclusion. As mentioned, the belief set (b)–(d) does not necessitate (e); whether or not she decides to avoid the piece of fudge depends on her relevant desires (i.e., her desire to eat fudge or to avoid unappetizing food). In the event that the subject does approach the piece of fudge, beliefs (b)–(d) can then be used to explain a residual sense of disgust or discomfort she may experience in eating the dog feces-piece of fudge. The second response is to note that (g) can be blocked, because (f) may not apply to the piece of fudge (see main discussion).

[11] In light of this discussion, it would be more accurate to attribute to her the belief that the piece of fudge is composed of a delicious but appealing (from purely a gustatory point of view) substance. The subject expresses reluctance to eat the dog feces-shaped fudge because she finds it to be visually unappealing. What this shows is that taste is not the sole consideration in a person's decision whether or not to consume some particular food item.

[12] This set of beliefs, as noted, may have to be adjusted slightly to explain some specific cases of belief-discordant behaviors. Two kinds of adjustments, in particular, deserve a mention. First, some belief-discordant behaviors may lack an affective component. To explain these cases, it would be sufficient simply to attribute to these subjects a belief that has to do with the apparent stimulus in question, and whatever (d)-like belief that connects the content of (b), to some behavioral disposition. The resulting explanation, in other words, would lack a (c)-like belief. Second, some cases of belief-behavior mismatches will require premises in addition to the aforementioned structure. For instance, Gendler discusses an example in which a person deliberately sets her watch five minutes fast as a strategy to better manage time (2008a, p. 640). To explain this case, it would be appropriate to attribute to her the following beliefs:

> (o) It looks like my watch is reporting that I am in a world where the time is tþ5.
> Therefore,
> (p) I am in a world where the time is tþ5.
> (q) If I am in a world where the time is tþ5, I am running late.
> (r) If I am running late, I should hustle.

This explanation differs from the earlier ones in that it contains an additional belief, namely, (p). Whether the subject hustles depends on whether she recognizes that the inference from (o) to (p) fails, in which case (q) and (r) will no longer apply. My view is that this case, contrary to Gendler's view, does require that the person temporarily forget that she has set her watch five minutes fast. It is worth noting that this case differs in an important way from the example of the dog feces-shaped piece of fudge. The reason the person ceases to hustle upon realizing (o) and _(p) is that (q) and (r) apply only if (p) is true. Once the person remembers that she is not in a world where the time is tþ5, she understands that she is not running as late as she thinks. In contrast, the subject in Rozin's experiment would still be reluctant to approach the dog feces-shaped object, despite knowing that it is in fact a piece of fudge, because (c) and (d) are not concerned with what the object is but with its appearance. These beliefs do not have to do with dog feces, but with things that look like dog feces. At any rate, regardless of the sort of adjustments that have to be made to explain these additional cases above, the resulting explanations still consist of a set of beliefs, one of which is responsive to an apparent stimulus, while the others are related to the content of the (b)-like belief.

[13] Here is an attempt to distinguish the cases which my argument covers from those that it does not: the former cases are ones in which the subject at least is explicitly aware that the stimulus in question is having some effect on her. For example, the subject in

Rozin's experiment is clearly aware that the stimulus is affecting her in some way (i.e., she is aware that the dog feces-shaped piece of fudge is somehow impacting her), whereas the same cannot be said of subjects in the sentence scrambling tasks. In fact, most of them do not even register that many of the words with which they are presented are associated with the elderly.

[14] Likewise with behaviors involving the frog and beebees, and the puppy and its mirror image.

[15] Here is perhaps an alternative way of making the argument. To retain Gendler's usage for the purpose of establishing a fixed reference point, I have referred throughout the paper to the circumscribed behaviors as ''belief-discordant behaviors.'' However, a consequence of my argument is to show that these cases do not turn out to be so perplexing or baffling after all. In other words, they do not quite count as instances of belief-discordant behaviors. To describe them as such would seem to place too much emphasis on one belief that is seen to be at variance with the subject's behavior (even though, strictly speaking, they could still be called belief-discordant behaviors). But as I have shown, there are also other plausible beliefs in place or at work to exert an influence. Once these beliefs are added in, the subject's behavior does not turn out to be as discordant as it first appears. Indeed, it may even be viewed as appropriate and rational, entirely consistent with certain aspects of her cognitive state. For example, the subject's reluctance to eat the dog feces-shaped fudge is undoubtedly discordant with her belief about fudge in general as an appealing and delicious substance, but it ceases to be so once other plausible beliefs come into play: beliefs about how that particular piece of fudge looks, how she feels about things resembling dog feces, and how she is generally disposed to behave toward disgusting things. In short, her reluctance to eat something that repels her seems appropriate, rational and concordant with her beliefs. Given that these cases are not as perplexing or baffling as Gendler supposes, they do not contribute to the overall argument in favor of the notion of alief.

References

Gendler, T. S. (2008a). Alief and belief. Journal of Philosophy, 105, 634–663.

Gendler, T. S. (2008b). Alief in action (and reaction). Mind & Language, 23, 552–585.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust:

Ascale sampling seven domains of disgust elicitors. Personality and Individual Differences, 16, 701–713.

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. Journal of Personality and Social Psychology, 50, 703–712.

Rozin, P., Haidt, J., & McCauley, C. R. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), Handbook of emotions (2nd ed., pp. 637–653). New York: Guilford Press.